**Rohan Gulve**

**Himalaya Patel**

**Rohith Pabbathi**

**Jeff-Ray Abanilla**

**Data Analysis on Cab Service Using R Programming**

**Introduction**

R is a language and environment for statistical computing and graphics. R provides a wide variety of statistical and graphical techniques in order to display the data in a meaningful and understandable way. It is designed to seamlessly incorporate compiled code which gives you all the benefits of a interactive language while allowing you to capitalize on the speed of compiled code.

We used R programming to analyze cab services based on the patterns and trends. The primary goal is to find the busiest time of the day and to use that information to provide a better overall service to customers during times of high load. A cab company can allocate a sufficient amount of resources to the area that will most likely have a large amount of customers and at the right time.

**Data Collection**

The raw data set was downloaded from the NYC Taxi and Limousine Commission. The website provided monthly trip sheet data in CSV format. Each file was more than 12GB in size and recorded all trips in the corresponding month.

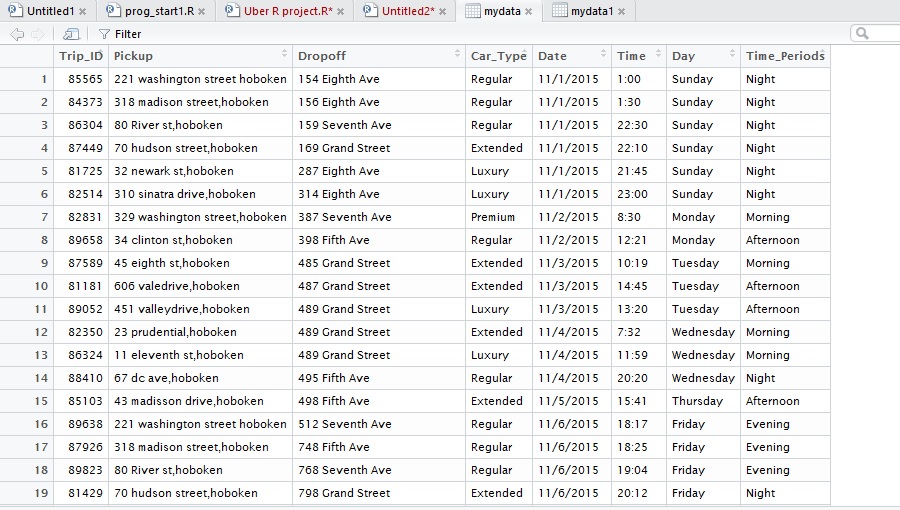
**Dataset Preparation**

The data set was too large to be entered into the R programming and we only needed specific columns for our project. The columns we focused on for this project were Transaction ID, Days, Booking time, Date, and Car Type. Pick-up and drop-off locations was also included in our data set but we were unable to include it into our project because the locations were very specific.

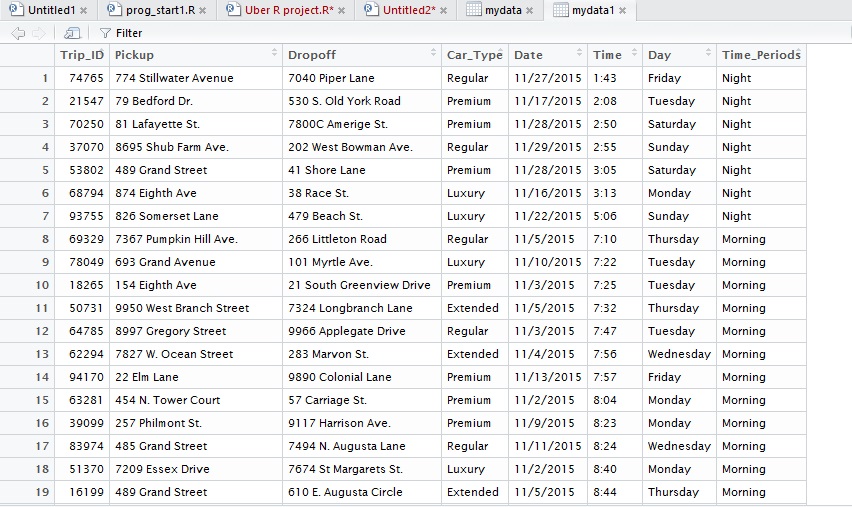
The CSV data still had too much data to be processed so we had to use a free program call CSV Splitter by ERD Concept 7 to split the main CSV file into sizes that can be handled. This program asks for how many rows you would like to limit the file size and breaks the main file into those files sizes.

Dataset Examples:

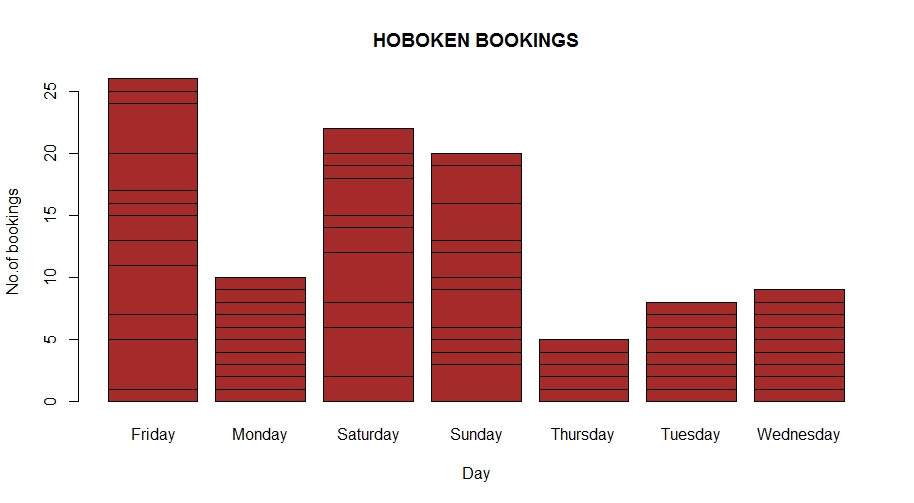
Hoboken



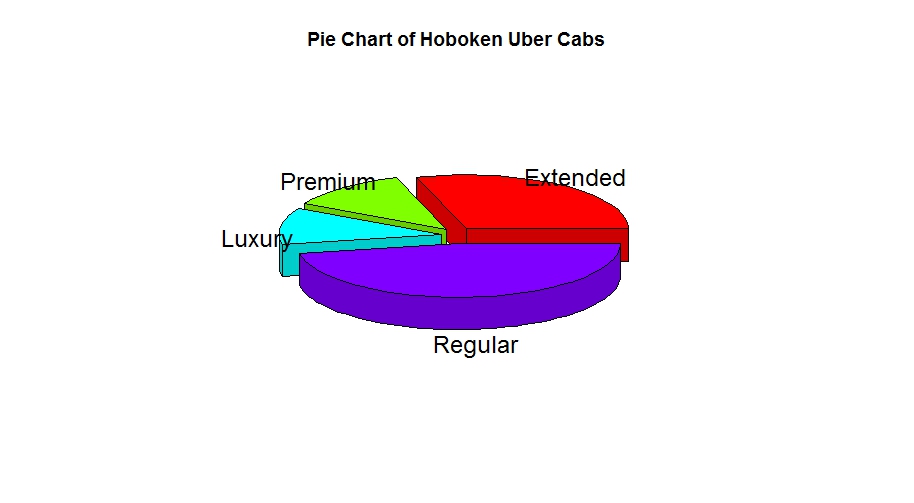
Manhattan



**Charts**



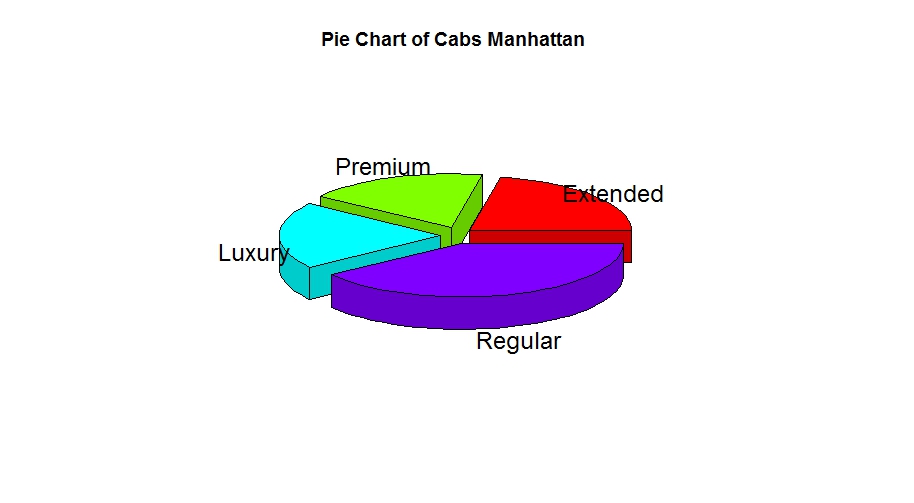
This chart shows the how many bookings in Hoboken on a daily basis. There are more bookings on the weekend as compared to the weekdays. Most people probably take their own cars instead of cabs because this isn't a heavily congested area.



This pie chart shows the breakdown of what type of transportation was booked in Hoboken. Regular cars were the majority, followed by extended, premium, and luxury.



Manhattan shows more bookings in the weekdays. This is probably because the area is so congested and is a corporate hub that it makes cabs a preferred transport. Limited parking and public transport from out of the city workers would make cabs the better option.



This pie chart shows what type of vehicles were booked in Manhattan. Regular cars were still the most booked, but there were more Premium and Luxury cars being booked when compared to Hoboken. Since this area is a corporate hub, there will be more people than can afford the higher costing options.

**Packages**

* ggplot2 – Maps variables to graphical aesthetics
* plotrix – Creates various plots
* readr – Allows CSV file to be imported and read by R Studio

**How our data and charts can make a difference**

This data can be used to analyze trends in where the customers will be when cabs are needed and what type of transported will most likely be requested.

**Conclusion**

Hoboken has more cab requests during the weekend and Manhattan has more requests during the weekdays. Due to the weekend break and nightlife, the number of requests in Hoboken were more on the weekend as compared to the weekdays. Manhattan has the opposite pattern of bookings than at Hoboken because it is a corporate hub and not as car friendly.

Using these conclusions, we can suggest several ways to improve services.

-Allocate more cabs to Hoboken during the weekend nights.

-Allocate more cabs to Manhattan during rush hour in weekdays.

-Provide discounts to users where cab request intensity is low.

-Fare hike to users where cab request intensity is high.

**References**

<http://www.erdconcepts.com/dbtoolbox.html>

<http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml>

<http://www2.elc.polyu.edu.hk/cill/reports.htm>

<https://docs.oracle.com/cd/E11882_01/server.112/e40540/toc.htm>